FRONTIERS

OF HEALTH SERVICES
MANAGEMENT

BRENT C. JAMES

Implementing Practice guidelines through Clinical Quality Improvement

Summary

The American health care delivery environment is changing. As provider-at-risk payment strategies become increasingly dominant, they will force health care providers to replace old strategies that measured and managed revenues with new strategies that measure and manage costs. Quality improvement (QI) theory provides a set of tools to do exactly that to understand, measure, and manage health care delivery processes and their associated costs. As a methodology for process management, QI theory merges case management, practice guidelines, and outcomes research into a single coordinated effort. It appropriately redirects management focus to care delivery processes, rather than to physicians. It also defines and illustrates a set of principles by which health care administrators can constructively team with physicians to find and document the best patient care outcomes at the lowest necessary cost, using QI-based practice guidelines as a decision support and measurement tool.

BRENT C. JAMES, M.D., is Executive Director, IHC Institute for Health Care Delivery Research, Intermountain Health Care, Salt Lake City, Utah.

American health care is changing. Ten years ago most American hospitals worked under a cost-plus system. Long-term financial survival required that a hospital's leaders manage *revenues:* They had to be certain that patients came to their facility to receive care (they tried to increase utilization) and they had to ensure that they added the right amount of financial margin to each service they provided. The natural unit of analysis in such an environment was a department. Many hospitals built sophisticated computer systems to *track* financial data at a departmental level, and used those systems as their primary source of management information and decisions. The medical staff bore responsibility for clinical quality, largely independent of administrators. Physicians also controlled the flow of patients into most hospitals. Therefore, hospital administrators often treated their medical staffs as their primary customers.

Then came 1983. That's the year the federal government first began to implement the Diagnosis-Related Group (DRG) Prospective Payment System (PPS). Suddenly, for about 30 percent of a typical hospital's case load, the size of the cost-plus margin no longer mattered. The government paid a flat rate per case regardless of the hospital's charges or costs. Hospitals initially shifted revenue shortfalls from government programs to other health care payer segments (a major factor in apparent hospital price inflation (Dranove, Shanley, and White 1991)). In response, many large private purchasers began to develop their own "provider-at-risk" strategies (i.e., managed care – per capita or per case payment) to limit health care expenses. Under those new structures, health care purchasers and third party payers began to supplant physicians' control of patient flow. Payers and patients became the hospitals' primary customers.

The medical staff bore responsibility for clinical quality, largely independent of administrators.

As the decade progressed hospitals saw larger and larger proportions of their patient volume shift to the provider-at-risk column. The phenomenon is most prominent on the west coast, where some community hospitals currently report that more than 90 percent of their inpatient volume comes through managed care contracts. It is gradually sweeping toward the east coast, engulfing localized pockets of heavy activity (such as the Minneapolis/St. Paul area) as it goes. And the trend is accelerating. For example, strategic planners at Intermountain Health Care (IHC–a 24 hospital system in Utah, Idaho, and Wyoming) initially estimated that provider-at-risk contracts would

increase from their current 60 percent penetration to about 85 percent of the system's total inpatient volume by the turn of the century. But vigorous political efforts to control health care costs, at both a national and local level, may reduce the time required to reach that level of managed care penetration to just three or four years.

As the provider-at-risk environment grows, hospitals are trying to use accounting adjustments to adapt their old revenue-based financial systems to the new reality. "Contractual allowance" or "deductions from revenue" measures what hospitals are *not* paid, relative to their charges. If managed care contracts account for most of a hospital's business, administrators can easily set their contractual allowance to any desired level by adjusting the hospital's charges, without affecting their net (real) revenues. In a provider-at-risk environment hospitals can no longer guarantee their long-term financial survival by managing revenues. Revenues are a fixed value, established through highly competitive, price-sensitive contract negotiations. Survival lies on the other side of the financial equation: Hospitals must begin to manage *costs.*

PROCESS MANAGEMENT

In a provider-at-risk, cost-based environment the natural unit of analysis and management is *a process,* not a department. A process is a series of linked steps, often (but not necessarily) sequential, designed to *cause* some set of outcomes to occur. The idea of a process not only aptly describes health care delivery but any repetitive human activity designed to add value, transform inputs into outputs, or cause some set of specified outcomes to occur. Processes usually span departments and facilities. Failures, which damage quality and increase costs, usually cluster around the interfaces, where one department or group hands off to another in the course of a single process.

Start with the idea of a process. Add to it fundamental knowledge of systems (processes interacting together), basic human psychology in a work setting, variation (statistics), and a theory for systematically acquiring and applying new knowledge (Deming 1990; Berwick 1993). Build a method to efficiently, effectively manage processes over time. The end result is the *methodology* (as opposed to the complementary philosophy) of quality improvement theory. In fact, implementing a total quality management strategy can be viewed as systematically redesigning a health system's infrastructure-its clinical data systems, financial data systems, human resources (policies and training), and culture-so that it is possible to manage and improve health care processes within a provider-at-risk environment.

Process management is also a common thread that brings together a number of current national health care initiatives: When focused on clinical processes of care, process management is case management. It blurs the line between operations and research, and provides a direct link between health care delivery systems and outcomes research (James, Horn, and Stephenson 1992). Finally, practice guidelines are explicit descriptions of preferred clinical processes. From that viewpoint practice guidelines are just a form of process management. Clinical quality improvement methods supply a set of tools to iteratively implement and modify such practice guidelines.

In 1989, the Congress of the United States formed the Agency for Health Care Policy and Research (AHCPR). Within its enabling legislation AHCPR has two specific missions: It must initiate studies to measure the outcomes of common health care interventions, and it must generate practice guidelines that codify research and consensus findings regarding best health care practices. These activities were mandated in the (untested) belief that they could eliminate inappropriate medical interventions and reduce health care costs (Institute of Medicine 1990). In addition to AHCPR, many professional groups, health care purchasers, and commercial enterprises are working to generate practice guidelines-often, though, with different objectives, different definitions, different levels of sophistication, and unequal quality in their final products.

Practice guidelines are explicit descriptions of preferred clinical processes.

It is hard to generate good guidelines. Special literature review methods, formal consensus techniques, sophisticated meta-analyses, and a very significant amount of effort are usually required. But many hospitals have the same hopes that prompted the U.S. Congress to launch AHCPR: They believe that, if they are to survive in the growing provider-at-risk environment, they must manage costs. They believe that practice guidelines may help them to document better patient care while controlling costs. Under many different names, hospitals therefore are trying to implement practice guidelines to manage health care delivery, whether they generate the guidelines themselves or obtain them from a third party.

The purpose of this article is to explore practical issues surrounding the physicians' role in the *implementation* of practice guidelines in American hospitals. It leaves the *generation* of practice guidelines to other sources (Institute of Medicine 1990; Eddy 1992). It also largely ignores the practical aspects of information systems and

organizational structures to handle the data associated with guideline implementation. With regard to physicians and guideline implementation, this article makes three key arguments to two crucial groups:

> **Physicians**: It is more important that you do it the same than that you do it "right."
>
> **Administrators:** It is more important how you implement than what you implement. The aim is to manage clinical processes, not to manage physicians.

To that end, this article first reviews several important quality improvement concepts central to any discussion that touches on cost, quality, and practice management. It then defines practice guidelines and introduces tools to document and manage them in a practical setting. It next presents a real-world case study in which a practice guideline was used to successfully manage a care process, simultaneously reducing costs and improving patient outcomes. Finally, it draws from the case study to discuss practical issues related to generating and implementing practice guidelines.

This article does not distinguish between *protocols* and *guidelines*-it uses the terms interchangeably. Many presently available guidelines (e.g., those published by AHCPR) lack sufficient detail to allow direct implementation. A potential user must first add a level of detail and definition that allows specific practice recommendations and measurement. With that level of detail a guideline becomes a protocol. But in terms of the physician relationships discussed here, the distinction is not critical.

Because of the physician focus of this article, all of the discussions and examples it uses center on clinical care delivery. But exactly the same techniques apply to nonclinical (administrative) support processes. Very often key clinical processes succeed or fail depending on the quality of the support processes upon which they rest. Further, it is not uncommon to find cost savings within administrative support processes that match or exceed those found in clinical processes. As a methodology, quality improvement and practice guidelines apply just as well to a hospital's administrators as to its clinicians.

IMPLEMENTING PRACTICE GUIDELINES: BACKGROUND PRINCIPLES

This article assumes that the reader is familiar with, and understands the implications of, several central principles of quality improvement theory. Those background principles are briefly outlined here, with

references that provide detailed discussions of their rationale and characteristics.

- *Quality controls costs.* Stated more accurately, quality and cost are two sides of the same coin. They are so tightly intertwined that it is impossible to act on one without acting upon the other (James 1989). Quality interacts with cost through three explicit mechanisms; Quality waste (costs fall as quality improves), productivity/efficiency (costs fall as quality holds stable), and cost-effectiveness (quality improves but at a higher cost). Quality also affects costs through secondary mechanisms such as the costs of attracting new customers, warranty (malpractice) costs, employee replacement costs, the costs of low employee morale, and long-term effects of low quality on an organization's standing within a community (the "ripple effect").

  Quality waste alone accounts for 25 to 40 percent of all hospital operating costs (Anderson and Daigh 1991). It is a particularly useful concept: As quality improves, it *causes* costs to fall a very favorable combination for attracting new patients. The idea of quality waste also provides a means (by seeking waste and rework) to identify areas for improvement.
- People show a consistent, predictable response when confronted with data that purport to document substandard performance. Scherkenbach (1991) called that reaction *The Cycle of Fear* (James 1992). The first phase of the Cycle is *denial* ("kill the messenger" or shift the blame)-"my patients are sicker." During the second phase, those individuals being measured begin to *filter the data* ("game the system") - as they generate the data that others will later use to evaluate them, they change how they assess and record that information so as to cast themselves in a more favorable light. The final phase is *micromanagement*-they know that they are outliers, but have no idea how they came to such a position. They therefore try anything they can imagine that might improve their apparent performance. Even though the vast majority of such efforts are ineffectual, and though some may do active harm, at least they are showing a good effort.

  Many quality assurance programs bog down in the Cycle of Fear, devolving into meaningless measurement and reporting. In such situations the aim is to meet regulatory requirements-a subtly different goal than patient care improvement. Those being measured soon become cynical.

They lose all faith that the process can ever produce better patient care. More importantly, most of the energy expended on quality goes to arguments about methodology, rather than improvement.

- Traditional quality assurance uses thresholds (standards) to define acceptable and unacceptable performance. Providers who do not fall below the standard are judged "good enough" - their quality is acceptable. Most providers approach quality from that viewpoint, in order to pass regulatory review. But often it is possible to perform at levels far better than a standard that defines lowest acceptable performance. Every time a provider fails to be the best they can be, they harm their patients and they waste money (through quality's relationship to cost). Within health care *"good enough" is never good* enough-the only acceptable goal is to find, implement, and consistently perform the best possible care processes (James 1992).
- Finally, *"find and implement the best" is a more effective strategy than "find and eliminate the worst"* to improve patient outcomes and reduce costs (James 1992). It's not that traditional quality assurance doesn't work; it just doesn't work very well when compared to quality improvement. Two principles are involved in quality improvement: Process operators use measurement tools to (1) eliminate *inappropriate* variation (usually in care process steps) then *(2) document* continuous improvement (usually in outcomes). "Find and implement the best" redirects energy from finding fault (and the natural defensive response that it provokes) to finding solutions. It creates a much more positive atmosphere within which to measure, criticize, and improve health care processes.

IMPLEMENTING PRACTICE GUIDELINES: DEFINITIONS

In 1989, AHCPR commissioned the Institute of Medicine (IOM) to assess practice guideline development and evaluation. The IOM's 1990 report defined practice guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." The report catalogs many of the conflicting definitions of practice guidelines found in the medical literature and in practice, and attempts to distinguish among them. Eddy (1992) adds an important distinction to the IOM's

definition of practice guidelines. He defines *practice policies* as "preferred recommendations issued for the purpose of influencing decisions about health interventions." In contrast, *performance policies* (or clinical algorithms) "guide or review the performance of interventions, without concern for whether the intervention should have been performed in the first place." Practice policies describe "doing the right thing"-clinical indications that lead to a decision to apply a particular medical test or treatment. Performance policies define "doing it the right way." They describe the manner in which the test or treatment should be executed. Eddy further distinguishes three levels of practice policies, depending upon the degree of professional certainty about the outcomes of a particular practice and patients' preferences for the practice's predictable results: *Standards* describe practices with well-documented outcomes and virtual unanimity among patients about their desirability. A standard is a relatively strict rule that embodies a "best" clinical decision in essentially all circumstances. On that basis, deviations from standards should be rare. *Guidelines* apply to clinical interventions that have well-documented outcomes, but whose outcomes are not clearly desirable to all patients. They therefore should be followed in most cases, but must be modified for individual patients. Deviations from guidelines may be relatively common, as dictated by differences in individual patient circumstances. *Options* describe medical interventions for which outcomes are not known, patient preferences are not known, or about which patients are indifferent. Options are neutral with respect to recommending a particular medical intervention-they simply provide a list of credible choices.

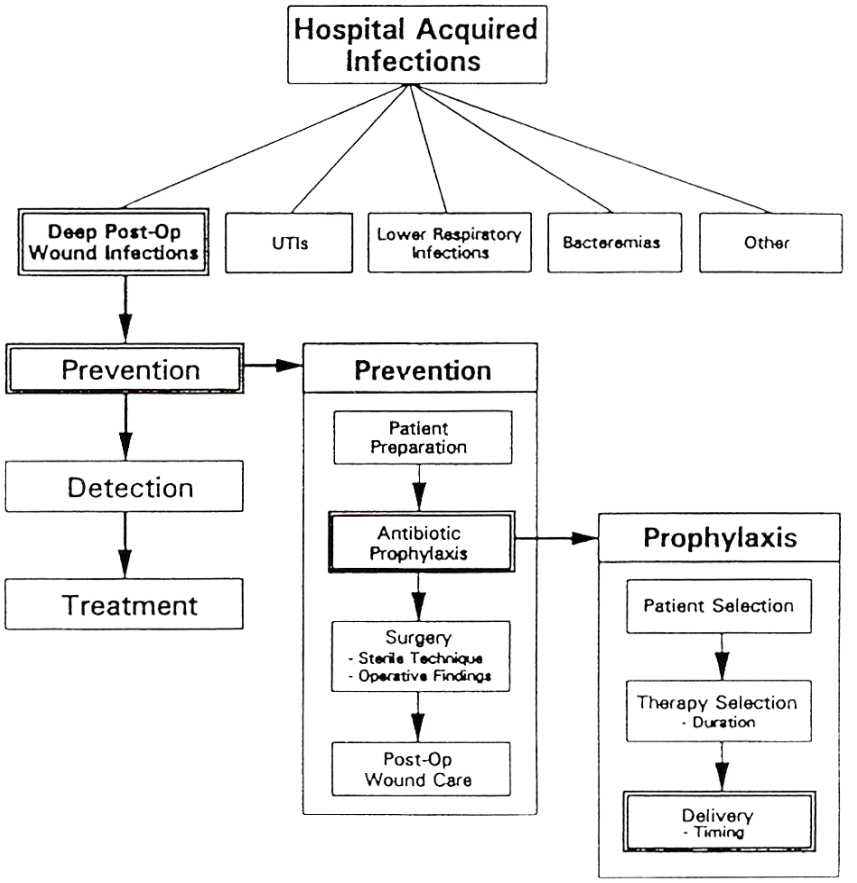
Guidelines apply to clinical interventions that have well-documented outcomes, but whose outcomes are not clearly desirable to all patients.

Practice policies and performance policies are obviously tightly interlinked. Eddy (1992) notes that correct performance of an intervention is immaterial if that intervention is not appropriate. But patients and clinicians choose a particular practice based on its documented outcomes. Those outcomes cannot be accurately established if the practice is not performed correctly. Seen another way, the "decision to intervene" and the "performance of the intervention" are sequential steps within a single process, with patient outcomes completing a feedback loop to inform the initial decision. Guideline implementation centers on performance policies, but draws from practice policies for source material.

Flow charts are a generic tool to document processes (several medical publications have recommended flow charts as a professional standard for recording, comparing, discussing and systematically improving practice guidelines)(Society for Medical Decision Making Committee on Standardization of Clinical Algorithms 1992; Pearson et al. 1992; Hadom, McCormick, and Diokno 1992). When a guideline is laid out as a flow chart, an important feature found in all processes becomes evident: Processes are inherently hierarchical. In other words, every box in a flow chart hides another, more detailed, sub-flow chart.

To illustrate, Figure 1 shows the process used at IHC's LDS Hospital (in Salt Lake City, Utah) to manage deep post-operative

Figure 1. Hierarchical Process of Care



A flow chart describing a hierarchical process of care used to manage deep post-operative wound infections. Such flow charts can be used both to control complexity when describing a clinical process, and to identify appropriate measurement points to monitor critical care processes. In some instances the simple act of documenting a process flow will improve coordination and reduce variation among care providers.

wound infections. Each step in the highest level process (Prevention → Detection → Treatment) is the outcome of a subprocess, with its own series of steps. Similarly, each step in every subprocess is the outcome of an even more detailed sub-subprocess, and so on down to an arbitrary level of detail. That means that every box in a flow chart is both a process step (of its superior process) and an outcome (of its subprocess). The terms "outcome" and "process step" are interchangeable, depending on the level of abstraction employed to manage -a particular subprocess at a particular point in time. Hierarchical flow charts are useful tools to focus attention and manage complexity. Starting with a high-level flow chart, detail is added within a focused area by expanding appropriate subprocess flow charts. When detail is no longer needed, the subprocess flow charts are collapsed back into their superior process steps. Flow charts play another essential role: They are the foundation for effective communications and systematic improvement within a clinical team. Without a written paradigm, differences in mental models, perceptions, and terminology make it extremely difficult to even discuss a complicated care process. With a written model guideline team members can identify differences in practice style, criticize specific steps in the model, and recommend improvements. Finally, when monitoring a clinical process, a flow chart identifies measurement points. It shows the data that are needed to track both performance and outcomes within a particular process.

Without a written paradigm, differences in mental models, perceptions, and terminology make it extremely difficult to even discuss a complicated care process.

Decisions (practice policies) and execution (performance policies) are embedded throughout a clinical flow chart. For example, the use of Antibiotic Prophylaxis is the second performance step in the Prevention subprocess for deep post-operative wound infections. But the first step in the Prophylaxis subprocess is Patient Selection. That step requires a decision. Clinical indications (a practice policy) identifies patients who should receive prophylaxis. Collecting and evaluating the information necessary to decide whether to use prophylaxis for a particular patient is a process in itself, with its own decision and performance steps. Nearly every performance step depends on underlying decisions, while nearly every decision step depends on underlying performance.

Eddy (1992; 1990-1992 (series)) has championed the use of meta-analytic methods to extract and synthesize important scientific information to guide clinical decisions. His work provides a critical

service in those instances when scientific data are available. Leape, Kosecoff, Chassin, Brook and other researchers at RAND have developed formal consensus techniques for use when scientific data are not available (Park et al. 1986). But given a practice guideline that describes a performance policy, based on appropriate practice policies, how can a hospital work with a clinical team to manage the process and document (for patients, purchasers, and regulators) that effective, efficient care results?

IMPLEMENTING PRACTICE GUIDELINES: A CASE STUDY

Adult Respiratory Distress Syndrome (ARDS) is a disease of the lungs. It often appears as a complication of an underlying pneumonia or shock and multi-organ failure. For example, one common precipitating cause is a simple viral pneumonia-a chest cold, of the sort that many individuals experience during the winter. For reasons that are not clearly understood, some patients' lungs react to the pneumonia by secreting fluid into their air spaces. As the lung's air spaces fill with fluid the lungs are not able to move oxygen into the blood (hypoxemia). The fluid also makes the lungs stiff (noncompliant), difficult to inflate and deflate as the patient breathes. Traditional treatment depends on mechanical respiration (ventilator support), with high oxygen concentrations and constant high air pressure to force oxygen into the blood despite the fluid. If the patient remains alive until the underlying pneumonia or shock resolves then their ARDS will often clear. Those who live usually achieve a complete recovery and regain normal health.

While ARDS affects both genders and all ages, it concentrates mainly within young men, in their twenties and thirties. Each year it accounts for about 15,000 cases in the United States. Historically, among all patients who developed ARDS only about one third survived.

During the mid-1970s pulmonary researchers developed an alternative therapy to use in place of stand-alone ventilator support. Called extra-corporeal membrane oxygenation (ECMO), it used a heart-lung machine, connected through the patient's femoral artery and vein, to oxygenate the patient's blood outside their body. In theory, that would keep the patient alive until their underlying pneumonia or shock resolved and their ARDS cleared. The researchers also established metrics (the ECMO criteria) that identified a subgroup of ARDS patients who were at a particularly high risk for death. Historically, about ten percent of patients who met the ECMO criteria survived. Clinical trials conducted on patients who met the ECMO criteria eventually demonstrated that ECMO was no better in

preventing ARDS deaths than standard ventilator therapy. The ECMO therapy was therefore abandoned. But pulmonary researchers have continued to identify and track ARDS patients who meet ECMO criteria. Recent estimates of survival among ARDS patients who meet ECMO criteria reach as high as 15 percent.

In the 1980s an Italian pulmonary research group reported a variant of ECMO that they claimed significantly improved survival in ARDS patients. In addition to oxygenating a patient's blood outside the body, they added equipment to simultaneously remove CO_2 and other waste products (extra-corporeal CO_2 removal, or $ECCO_2R$). A pulmonary research team at IDS Hospital received a grant from the National Heart, Lung, and Blood Institute to test the new therapy in the United States. They planned a randomized clinical trial to compare $ECCO_2R$ (the treatment arm of the trial) to standard mechanical ventilation (the control arm) for ARDS patients who met ECMO criteria.

A key factor in clinical trial design centers around the idea of consistency. If a clinical trial is to accurately compare two competing treatments, then each of the treatments must be applied in a consistent fashion. Otherwise it is impossible to judge whether differences in patient outcomes are due to the treatments or to variations in their application. Trials therefore usually use protocols to describe, in detail, the manner in which each treatment will be delivered. As the LDS Hospital research team began to construct the $ECCO_2R$ clinical trial they had an important insight:: They recognized that, despite the fact that they had practiced together for many years, cross-covering each other on the same patients, they didn't manage ventilators in a consistent fashion. Those differences went beyond variation among the physicians, nurses, and therapists in the group. Individual clinicians showed differences in practice patterns from patient to patient. In fact, it appeared that a single clinician sometimes was inconsistent when treating the same patient from day to day, or even from morning to evening rounds.

If a clinical trial is to accurately compare two competing treatments, then each of the treatments must be applied in a consistent fashion.

The team therefore decided to generate a detailed protocol-a practice guideline-to oversee ventilator management on the control arm of their clinical trial. They first performed a careful literature review to identify important research findings that should guide their decisions and care practices. They then used formal consensus

techniques to fill in those parts of the guideline not covered by the scientific literature, and represented their new guideline as a flow chart. It described ventilator management for ARDS in detail, being more than 35 pages long with an average of more than twenty major decision nodes per page.

But when they had completed the guideline they had a second important insight: Much of their new ventilator management guideline was based upon consensus-"expert" opinion, generated as a theoretical exercise far from the treatment of real patients. They had no *data* to demonstrate that the consensus portions of their guideline were correct. Because of the consensus process, they had no scientific basis to argue that the guideline represented best care for real patients.

The ECCO2R team therefore chose to use their new ventilator guideline in a very innovative way. They reviewed the guideline with all involved clinicians (nurses, physicians, and other allied health professionals) so' that everyone understood its content. They built a measurement system to track whether the clinicians followed the guideline's recommendations, at the level of each detailed decision covered in the document. Finally, they placed a copy of the guideline at the bedside of every ARDS patient being treated with a mechanical ventilator, and asked that the clinicians follow it. But if a clinician disagreed with a guideline recommendation, the team instructed the clinicians to follow their own judgement, not the guideline. In such circumstances they assumed that the guideline was probably wrong, not the clinician. After all, they knew and trusted the clinicians on the team. It was the guideline that had yet to prove itself with demonstrated results.

The research team then carried their reasoning to the next logical step: If a clinician failed to follow the guideline, leading to the assumption that the guideline was probably wrong for that particular decision, then they had an opportunity to correct the guideline. They therefore automatically added that clinical case and the associated guideline-based decision to the agenda for their next weekly staff meeting. That meant that they were able to discuss each questionable guideline recommendation as a group, in the context of a real case. In those meetings they stripped identifying information from the cases in order to avoid the Cycle of Fear. Their aim was to fix the system, not fix blame. They wanted to agree upon a "best" treatment processes as a group, not single individual team members out for criticism.

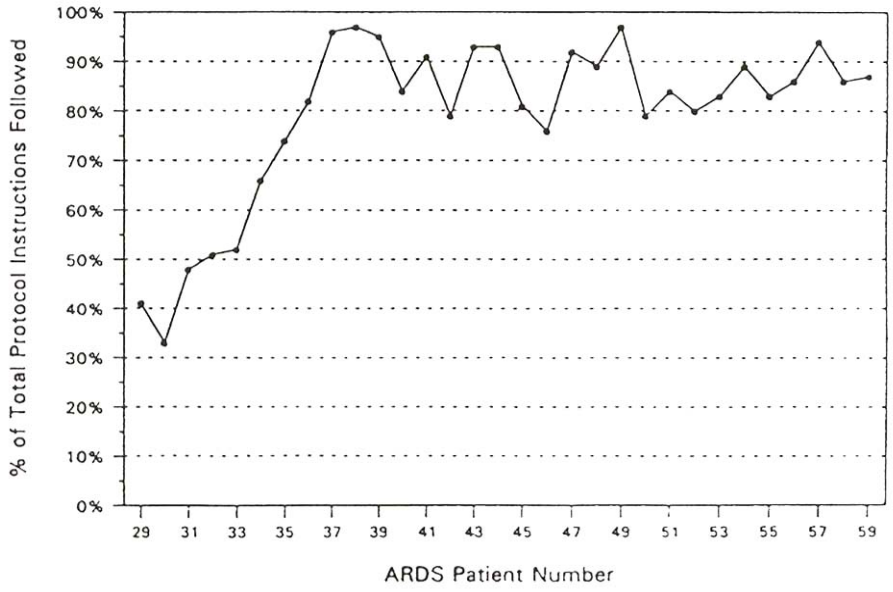
Three possible courses of action are possible in such a setting:

1. As the team examined a guideline recommendation, they could conclude that the guideline was wrong and change it.

2. The team could agree that the guideline was right. That sends a message not only to the clinician who had made the original decision, but to all members of the team, concerning their group consensus about best patient care, as codified in the guideline.
3. They could decide that the case was an outlier for that particular decision. No guideline can reasonably cover all patient variants.

Figure 2 shows guideline compliance rates as the team used their iterative review process (Henderson et al. 1990). Over a period of about four months, guideline compliance increased from under 40 percent to more than 90 percent (Henderson et al. 1992; Henderson et al. 1990; East et al. 1992a). In the early stages changes to the guideline were common. Note that the team never achieved perfect

Figure 2. **Ventilator Protocol Compliance (9/14/88–1/20/90)**

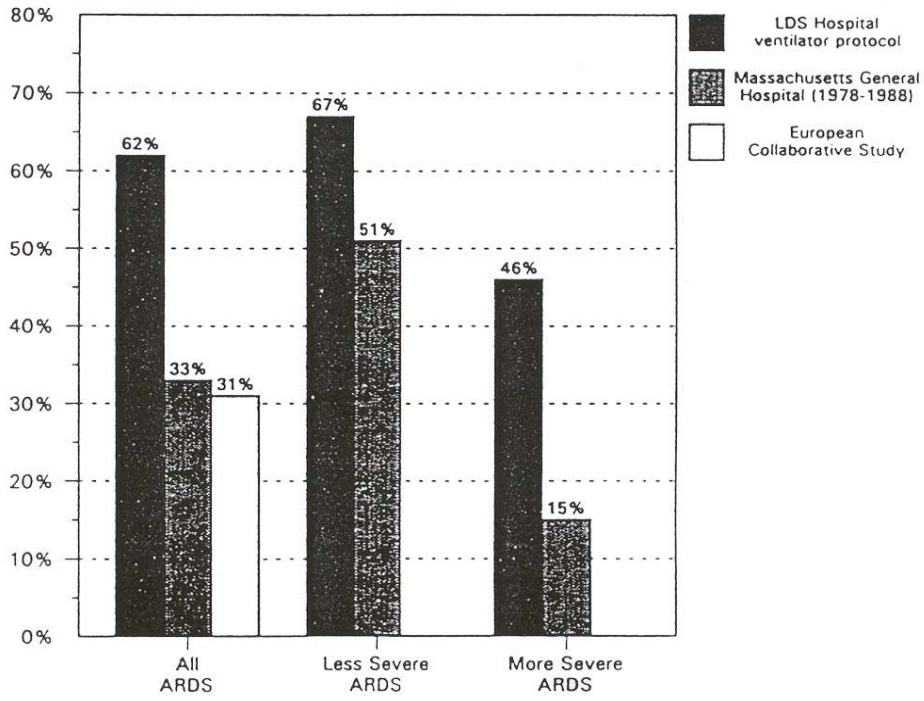


Percentage of protocol-based recommendations followed for Adult Respiratory Distress Syndrome (ARDS) patients, starting with the first patient (Patient Number 29, admitted to LDS Hospital's Pulmonary ICU on August 14, 1988) for whom the protocol was used through 30 consecutive patients (Patient Number 59, admitted on January 1, 1990). A typical treatment episode involved more than 200 protocol-based treatment recommendations. Roughly four months elapsed between Patient Number 29 and Patient Number 37. From Patient Number 37 on, most protocol noncompliances occurred either (1) when a patient was removed from the ICU (usually for either surgery or imaging), (2) as further improvements to the protocol were tested, or (3) as a consequence of the fact that few protocols are perfect (nearly all guidelines show some level of random noncompliance as clinicians address patient factors not anticipated by the protocol or factors that are so rare as to not justify inclusion in the protocol).

guideline compliance. No guideline will ever perfectly match every patient, or supplant clinical judgement (East 1992b). Statistical process control (SPC) provides an ideal tool to track noncompliances in a process, separating treatment deviations arising from differences in patient presentation (appropriate, common, or random variation) from those arising from external practice patterns that intrude into the treatment process (inappropriate, special, or assignable variation) (Ryan 1989).

Upon completion of the randomized clinical trial, $ECCO_2R$ achieved 38 percent survival for patients who met ECMO criteria. Stabilized ventilator management (as produced by the ventilator protocol) achieved 44 percent survival for the same patient group, better than the $ECCO_2R$ treatment arm and much better than the 9 to 15 percent survival expected for ventilator management from historical experience (Morris et al. 1992; Morris 1992). Figure 3 compares ARDS survival experience from several pulmonary research groups, covering cases beyond the LDS Hospital $ECCO_2R$ clinical trial (Zapol et al. 1991; Artigas et al. 1991).

Figure 3. Percent Survival for ARDS Patients



Survival among Adult Respiratory Distress Syndrome patients by risk class, comparing LDS Hospital's experience with stabilized ventilator management (a detailed practice protocol) with that of other major pulmonary research groups who did not use the ventilator protocol.

The team's experience with a practice guideline produced other interesting results:

- Physician time to manage these complex cases fell. That was because common, day-to-day decisions were pushed down into the system, where physicians did not have to consider them one case at a time. It wasn't that the physicians didn't think about the patient care issues involved, but that they addressed them for groups of patients, instead of case by case. That freed physicians to deal with the patients' interesting problems, that required a physician's oversight, or allowed the physician to see other patients. It also made the members of the team (physicians, nurses, and technicians) more predictable to one another, which may reduce friction and improve efficiency.
- If a patient lived, they may have left the intensive care unit (ICU) faster than similar patients had before the introduction of the ventilator protocol. That is probably because the patients could advance on the protocol 24 hours per day, rather than waiting for a physician to come on rounds and change orders.
- Stabilized ventilator management cost about $120,000 per patient who lived. $ECCO_2R$ (the next best therapy) cost more than $160,000 per patient who lived, not counting physician fees.

The LSD Hospital pulmonary research team is now supervising a follow-up randomized clinical trial that compares traditional ventilator management-for ARDS patients to stabilized ventilator management as produced by their ventilator protocol.

IMPLEMENTING PRACTICE GUIDELINES: LESSONS LEARNED

Lesson 1: The core problem is variation in clinical practice.

When members of the LDS Hospital ARDS research team recognized the variability of their own ventilator management practices, they built upon a long line of studies that demonstrate variation in medical practices. Glover first measured differences in the rates of tonsillectomy among various regions of England, beyond what could be explained by population differences, in 1938. Lewis provided additional evidence of the phenomenon in the United States in the late 1960s.

During the 1970s and 1980s, Wennberg formalized and extended analytic techniques for examining differences in surgical procedure use rates or hospitalization rates among communities. He called those methods small area variation analysis (SAVA). His studies again demonstrated that hospital admissions for some surgical procedures and medical diagnoses occurred at a much higher rate in some communities than other, similar communities, even after controlling for underlying population factors (Wennberg and Gittelsohn 1973; Wennberg 1985; Wennberg, Barnes, and Zubkoff 1982). Wennberg also showed that the range of inter-community variation was related to specific surgical procedures and medical diagnoses. When examining the rates of use for the same procedures and diagnoses in other countries, he found that some showed consistently low ranges of variation within all countries examined, while others showed consistently high ranges of variation within all countries examined. This was true even though the average use rates for each procedure or diagnosis varied significantly between the countries included in the study (McPherson et al. 1982).

The RAND team developed formal methods to generate measurable indications for several surgical procedures and medical hospitalizations.

The RAND team (Park et al. 1986) hypothesized that SAVA differences among communities could be explained by higher rates of inappropriate treatment in communities with high use rates. The RAND team developed formal methods to generate measurable indications for several surgical procedures and medical hospitalizations. For each condition they examined, they first performed a structured review of the medical literature. They then presented the resulting scientific information to a panel of expert physicians, drawn from the appropriate specialty area. Within each expert panel they used formal consensus techniques to derive extensive lists of appropriate, equivocal, or inappropriate indications for the treatment under study. Finally, they used their indications to measure the rates of appropriate versus inappropriate use of the targeted conditions in communities that showed high rates of utilization and communities that showed low rates of utilization for the procedure or hospitalization in question. They discovered that high use rates were not consistently associated with high rates of inappropriate indications (Leape et al. 1990). That is, geographic areas that showed low utilization rates for a particular surgical procedure or medical hospitalization often had as high a rate of inappropriate indications as other geographic areas that showed high

utilization rates for the same medical decision. They also demonstrated that inappropriate surgical procedure use and inappropriate hospitalization for medical conditions are surprisingly common, and that some procedures or hospitalizations show consistently high rates of inappropriate application, while other show consistently lower rates of inappropriate use (Brook et al. 1990; Winslow et al. 1988).

Wennberg's small area variation analysis and the RAND team's measures of appropriateness addressed a single class of issues: Both examined the decision to treat a patient, at the level of Eddy's practice policies (indications for treatment). A further set of studies (James et al. 1987, 1988) investigated variations in what happens to patients after they enter a hospital (at the level of Eddy's performance policies). Their Quality, Utilization, and Efficiency (QUE) studies tracked patients with comparable presenting disease, comorbidities, and outcomes hospitalized for transurethral prostatectomy (TURP), cholecystectomy, total hip arthroplasty, and permanent pacemaker implantation (Baird et al. 1988, 1989). Those studies showed that physicians used widely different amounts of specific care factors to treat similar patients, with differences among physicians ranging from 60 to 460 percent. Figure 4 illustrates two important process of care factors for TURP - true surgical procedure time and grams of prostatic tissue removed - among 16 urologic surgeons, across a group of comparable patients at four hospitals. Each factor varies by more than 200 percent across the physician group.

While well-designed studies document wide variations among physicians with regard to their decisions to apply treatment to patients and the manner in which those treatments are applied, anecdotal information suggests that practice variation may extend even further. Individual physicians appear to vary in how they diagnose and treat similar, sequential patients, beyond what would be expected from patient factors. Observation of very complicated patients (for example, ARDS patients on ventilators as described in the case study) suggests that physicians may vary from contact to contact, morning to night, in how they treat individual patients.

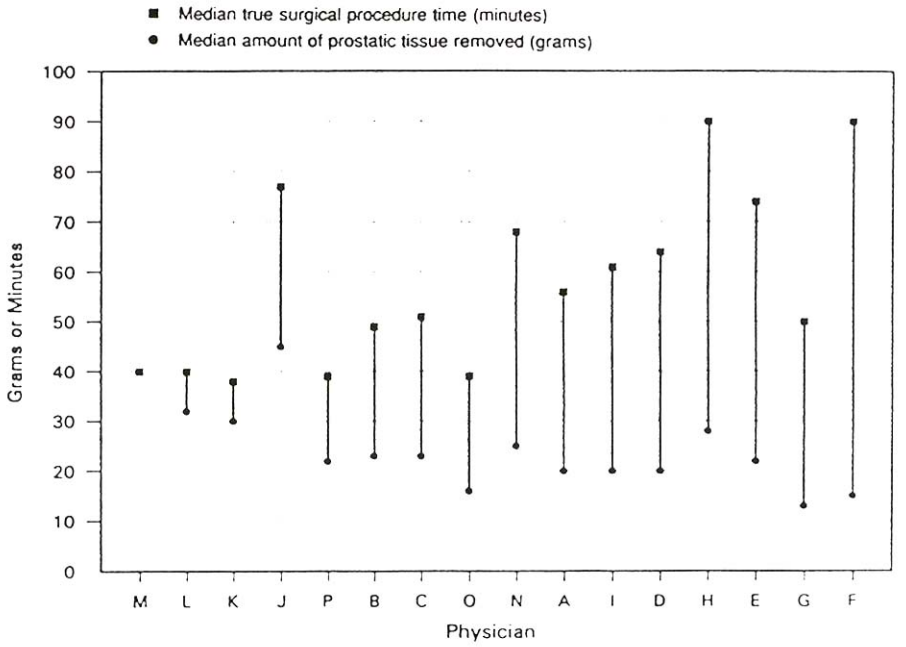
Eddy, Wennberg, and others have summarized possible causes of practice variation (Eddy 1984; Eddy and Billings 1988; Wennberg, Barnes, and Zubkoff 1982; James, Horn, and Stephenson 1992). The combined list is extensive, running to more than 60 different items. Interestingly, many of the most prominent causes are not under physicians' control. They arise from professional uncertainty:

- *80 to 90 percent of common medical practices have no basis in published scientific research.* In 1979, Williamson tracked

common medical practices for three subspecialties of internal medicine back to the medical literature. He estimated that fewer than 10 percent of the medical practices examined had any foundation in published research (Williamson, Alexander, and Miller 1968; Williamson, Goldschmidt, and Jillson 1979). Follow-up studies by the federal Office of Technology Assessment in 1985 (Bunker 1988; Institute of Medicine 1985), and the Office of Medical Applications of Research in 1990 (Ferguson 1991; Dubinsky and Ferguson 1990), generated estimates of 10 to 20 percent and less than 20 percent, respectively.

That does not mean that 80 to 90 percent of medical practices are wrong. They are based on a long history of medical tradition and experience, and probably help most

Figure 4. Variation in Transurethral Resection of the Prostate (TURP) Practice Patterns among 16 Urologists



Variation among urologic surgeons for two important process of care factors when performing a transurethral prostatectomy (TURP): true surgical procedure time and grams of prostatic tissue excised. The cases in the study were similar in terms of the presence and severity of comorbidities on admission to the hospital and in terms of medical outcomes (complications and therapeutic goals). The surgeons are shown in order of grams removed per minute of surgery. Both factors (surgery time and grams of tissue) showed more than two-fold variation across the physician group. The length of surgery had a strong statistical association with grams of tissue removed: The longer a surgeon's procedure time, the smaller amount of tissue removed.

patients. But it does mean that for most medical practice we do not know what is best. Practitioners can hold legitimate differences of opinion about best practices.

- *Much of the scientific research that does exist is not available to medical practitioners.* Williamson et al. also documented that even when scientific research regarding best medical practices does exist, its diffusion into actual medical practice is slow and uneven (Williamson, Alexander, and Miller 1968; Williamson, Goldschmidt, and Jillson 1979). Given the size, complexity, and lack of methodologic consistency of the medical literature, that finding is not unexpected. An effort to find, evaluate and synthesize appropriate scientific articles for a particular medical topic requires special expertise. Most practitioners lack the tools, resources, and time for such an undertaking. Williamson, Lincoln, and Turner (1991) recommended a set of formal methods (called information synthesis) for that purpose (Goldschmidt 1986). Eddy has published meta-analytic techniques, and produced and distributed computer software, to address the same issue (Eddy 1992; Eddy, Hasselblad, and Shacter 1992).
- *Even such limited scientific information as is available may overwhelm the capacity of the unaided human mind.* The human mind is limited in its capacity to synthesize complex information to optimize outcomes. When working with the ventilator guideline presented in the case study, Morris (1992) found that experienced physicians were not able to manage more than four concurrent variables to maximize patient outcomes. Unfortunately, a typical ARDS case presented more than 200 active variables. A physician's patient care decisions depended on which small subset of variables the physician chose to analyze. As different variables thrust themselves into the physician's attention at different points in time, the physician's practices changed. Thus, a physician could vary from morning to evening in managing the same patient. While ventilator management for ARDS patients in an ICU is admittedly complex, most patient care decisions involve far more than four variables.

American medical practice is based on the notion of a physician-patient relationship. That ideal asserts that best patient care occurs when an individual physician advises a patient on factors affecting their physical and mental well-being, available health care responses, and likely outcomes, so that the patient can make informed decisions about their own health. A major corollary is the assumption that an

individual physician can *subjectively* integrate hundreds of disparate factors to accurately advise patient decisions. At a minimum, a physician must correctly synthesize the patient's underlying disease processes, their individual physiologic response to each of those diseases, various treatment options, likely outcomes for each possible treatment, and the patient's personal values and preferences, in order to provide an accurate list of options from which the patient can chose. But that model is incompatible with knowledge of the limited capabilities of the unaided human mind.

Perhaps Eddy (1990, no. 4) said it best: "It is simply unrealistic to think that individuals can synthesize in their head scores of pieces of evidence, accurately estimate the outcomes of different options, and accurately judge the desirability of those outcomes for patients.... All confirm what would be expected from common sense: The complexity of modem medicine exceeds the inherent limitations of the unaided human mind."

- *Humans are inherently fallible information processors.* McDonald (1976) demonstrated that, regardless of training or intent, humans make errors when handling data. Some types of errors, such as digit transpositions when writing numbers or misplacement of decimal points, occur more frequently. An individual's error rate is affected by stress levels (e.g., lack of sleep), complexity, and whether that individual is operating within their domain of specific knowledge. Simple errors can introduce variation into patient care. Carefully designed, robust processes can catch and correct or reduce the effect of human errors when they do occur.
- *Differences in observation-measurement error-can lead to differences in assessment and differences in treatment among physicians.* Koran documented frequent differences among physicians in physical examination findings, interpretation of diagnostic procedures, diagnosis, recommended treatments, and evaluations of the quality of care (Koran 1975). He notes, "The physicians studied almost always disagreed at least once in 10 cases, and often disagreed more than once in five cases, whether they were eliciting physical signs, interpreting roentgenograms, electrocardiograms or electroencephalograms, making a diagnosis (from incomplete information), recommending a treatment or evaluating the quality of care. Disagreements of this magnitude, if characteristic of clinical practice in general, cannot safely be regarded as inconsequential."

Beyond differences in patient care that arise from professional uncertainty, human limitations, unequal allocation of health care resources, and variation in measurement, patients differ from one another and individually differ over time. Among other inconsistencies, they have different values, different preferences, different symptoms, different physiologic response to disease, and different ways of interacting with health care providers. All of these factors cause appropriate differences in practice patterns. Any clinical process management system must group patients together in a way that takes those differences into account.

Lesson 2: Real benefits accrue to patients, payers, and providers when *inappropriate* practice variations decline.

Quality improvement uses statistical process control to separate *assignable* from *random* variation. Random variation arises from differences in a process's inputs (i.e., differences in patient presentation) or the sum of many small variations in process steps that cannot be tracked to specific, preventable causes. It is a physical, measurable attribute, representing the random noise in any real process. Assignable variation arises from identifiable causes that can be tracked and eliminated. Statistical process control graphs the probability that variation in a specific process measurement arises from assignable, rather than random, causes. In quality improvement jargon, assignable variation represents *inappropriate* variation. Random variation is not only appropriate, but expected. When it does not appear something is probably wrong.

It is important to distinguish between the two types of variation because each requires a different management approach. With assignable variation, the aim is to track the outlier points to their root causes then eliminate them, so they never intrude in the process again. On the other hand, random variation is a physical attribute of the process and its inputs. To reduce random variation, a provider must design a new process (usually a variant of the old process, generated by changing specific process steps), then scientifically compare its performance against the old approach. Quality improvement theory calls such a test the Shewhart Cycle, and summarizes its steps as "PDCA": Plan a change, Do it in a small subgroup, Check its performance against prior outputs, then Act (discard the change or fully implement it). Traditional medicine calls the same approach a clinical trial.

Quality improvement theory defines *a stable process* as a process that shows only random variation over time, with no assignable variation. *Process capability is* the ability of a process to achieve

its stated goals. For example, a process designed to prevent infections has an absolute goal of no infections. If only two percent of cases develop infections, then the process is 98 percent capable-it achieves its goal 98 percent of the time.

Quality improvement theory notes that it is impossible to measure a process's true capability unless that process is stable. Otherwise, variation in performance can alter (for the worse) the apparent efficacy of the process (the actual outcomes of an uncontrolled process in a community setting is called effectiveness). That is exactly the same idea embodied in the treatment protocols that define each arm in a controlled clinical trial. Such protocols guarantee that the trial's treatments are consistent from case to case. Otherwise, it is impossible to tell whether differences in patient outcomes arise from true differences in the capabilities of the treatments, or just variation in how they were applied.

In other words, if a clinical process shows inappropriate variation, it is impossible to even measure its true outcomes, let alone apply the scientific method (clinical trials) to systematically improve. But American medicine is rife with assignable variation. Much of it arises from practice differences among physicians. But for 80 to 90 percent of common medical practices, physicians' assertions of what is "right" for their patients is just a matter of opinion.

Quality improvement theory notes that it is impossible to measure a process's true capacity unless that process is stable.

Hence the statement to physicians regarding their patient treatment practices: It is more important that you do it the same than that you do it "right." For when, as a group, physicians develop consistent practices based on the best scientific information and peer consensus, they can accurately measure patient outcomes and apply the scientific method to systematically improve. No matter where a group starts, iterative application of the scientific method, informed by comparisons with other professional groups, will eventually lead to documented best patient care. But without consistent care delivery practices it is not even possible to accurately measure outcomes, let alone systematically improve.

The costs associated with a health care process are just one more outcome of the process. As such, process management techniques apply to them just as well as to medical outcomes. This is the basis for the widely held but unproven view that efforts to eliminate variation will produce less costly as well as better care.

In an increasingly competitive medical marketplace, it is critically important to physicians and hospitals that they are able to *document* effective, efficient patient care, improve both quality and cost over time, and share the results with patients, purchasers, and regulators. In a very real sense, within a provider-at-risk environment a provider's financial success is tied directly to the provider's professional success. Both depend upon the provider's ability to measure and manage variation.

Lesson 3: For most physicians, financial rewards are secondary to good patient care. For that reason practice management efforts that emphasize patient care quality are much more successful, even for managing costs, than those that focus on costs alone.

The foregoing list of major sources of physician practice variation overlooks one oft-cited factor. Financial incentives clearly affect medical decision making (Wennberg, Barnes, and Zubkoff 1982; Eddy 1984; Hillman et al. 1992; Mitchell and Scott 1992; Mitchell and Sunshine 1992; Swedlow et al. 1992). But financial incentives exist within the broader context of professional uncertainty. When forced to choose between good patient outcomes (as supported by credible clinical data) and their own financial gain, most physicians consistently elect to maximize patient outcomes (consider, for example, Maine's experience with falling surgical rates following publication of outcomes data (Caper 1991)). More than that, physicians almost exclusively use the language of quality when they argue practice issues among themselves. Even financial arguments are usually couched in quality terms (Wennberg et al. 1977). By concentrating on quality of care (at the level of professional uncertainty) process management can align the moral weight of the entire medical profession with its goals, and control the context within which financial decisions take place.

Physicians' response to financial incentives may arise partly from nonphysicians' fixation on costs: Many practicing physicians perceive that health care administrators, regulators, and payers care *only* about reducing costs. If an administrator's ill-considered cost control efforts damage patient care, the patient and the physician are left to face the ethical and legal consequences alone. But as quality improvement theory (and related experience) clearly demonstrates, quality controls costs. One of the best ways to control costs is to manage quality. One of the biggest hurdles IHC faced in implementing clinical process management (in order to improve both quality and costs) was overcoming the distrust that years of monomaniacal cost control efforts had built among physicians. As IHC's administration

has shifted its emphasis to best patient outcomes (with secondary cost control in a quality improvement setting) physicians' willingness to collaborate on clinical process management has steadily increased.

Lesson 4: Guidelines are nothing new to American medicine.

When the LDS Hospital research team began to develop a protocol to control ventilator management for ARDS patients, they followed models that have seen continuous use in American medicine since the early 1900s. Physicians routinely use guidelines in daily practice (even though they often apply them subjectively) because it helps them deal with complex decisions and makes them more efficient (guidelines save time, as the LDS Hospital ARDS team so clearly demonstrated). As an extreme example, residents and interns routinely purchase, carry, and rely on books of medical guidelines specifically designed to quickly summarize the diagnosis and treatment of common conditions.

Lesson 5: "Control" is a central issue.

Why, then, would physicians resist the implementation of practice guidelines at a hospital level? One major reason is that they fear a loss of control. They see hospital-level practice guidelines as straitjackets that mandate decisions, fail to recognize the full complexity of real patient care, and eliminate clinical judgement. More than that, they perceive that control is wrested from them only so that administrators can control costs. The responsibility for bad patient outcomes still rests with the physicians, even after their control over patient care (and, hence, patient outcomes) is gone.

In contrast, medicine's traditional guidelines are decision support tools that recognize the need for clinical judgement. The LDS Hospital ARDS team addressed this issue by giving the clinical team control over the guideline. The use of statistical process control to measure guideline compliance, which inherently recognizes a range of appropriate (random) variation while still preventing inappropriate (assignable) variation, reinforces the critical role of clinical judgment.

Lesson 6: Implementing process management requires a partnership between physicians and administrators.

The idea of continuous improvement/process management is a central tenet of the medical profession. Every physician, upon entering the practice of medicine, ethically commits to examine the treatments they give to patients and the outcomes they achieve, with an aim to

improve their treatments for future patients. In medical school and residency training every physician also forcefully learns that they cannot trust subjective data-objective information and evaluation are essential to good treatment decisions. But for some reason, when physicians leave training and enter the practice of medicine, they begin to evaluate their treatments and outcomes subjectively, in their heads. Because of that subjectivity their practices often resemble a series of small, unplanned, uncontrolled human experiments, based on the last journal article the physician had time to scan or the last drug representative who visited the clinic. Obviously, that kind of medical practice the kind practiced by almost all American physicians–has no chance of generating viable information about best patient care.

The central question, then, is "Why?" if physicians know that objective information is critical to the practice of medicine, why do they base their practice of medicine on subjective evaluations? As we implemented clinical process management within IHC, we had the opportunity to ask that question of many community-based and academic physicians, as well as examine its meaning in-our own medical practices. We concluded that practicing physicians do not have the resources, the time, or the training to deal with the masses of data required for objective practice management. But data management is a well-established ability within health care organizations.

Effective guideline implementation requires a partnership. Physicians, working as a peer group, supply clinical leadership. They have the clinical understanding necessary to oversee the content and direction of clinical guidelines. They can meet (as a group) to discuss best patient care and to review guideline compliance. The hospital supplies staff support. Hospital staff collect, collate, and analyze the clinical data, and support the generation and maintenance of other guideline-related documents (such as flow charts).

Lesson 7: Local consensus is essential for implementing guidelines.

Because most clinical practices have no firm basis in published scientific research, those who generate practice guidelines are often forced to rely on expert consensus to complete their work. But even when generated through formal methods, expert consensus is an inexact tool. Different consensus groups have different goals and use different techniques. They often generate different, even conflicting, guidelines on the same topic (Kellie and Kelly 1991; Audet, Greenfield, and Field 1990; Leap et al. 1992). Within a single consensus panel the experts often disagree, and their assessments change when they apply guidelines generated in a theoretical setting to real patients (Park et al.

1989). Perhaps most troubling, physician experts show wide disagreements when asked to assess underlying probabilities that are essential to consensus judgments (Eddy 1984; O'Conner 1988). For example, Eddy (1992) asked thoracic surgeons, sponsored by a professional society, to assess the chance of a particular outcome for a well-defined group of patients within a specific time period after surgery. The outcome was an essential element to determine when the procedure was appropriate. The surgeons' assessments ranged from zero to 100 percent. In light of many similar examples, there is real doubt that such a thing as a "medical expert" truly exists.

As the LDS Hospital ARDS team recognized, expert consensus suffers from the same deficiency that produces practice variation in the first place: There are no data to show that consensus guidelines are correct. In such a setting, if the aim is to stabilize a care process then systematically improve, local consensus among the complete care delivery group is far more important than the consensus of an expert panel. An expert consensus panel can provide a jumping off point, to get a practice guideline started. But that expert consensus must translate into local consensus if the guideline is to modify physician practices.

Lesson 8: Effective guidelines require feedback on compliance and outcomes, using credible clinical data.

In 1989, Lomas et al. tracked the implementation of a practice guideline covering repeat cesarean sections in Canada. The guideline was developed and widely distributed by the major professional society that represented obstetricians in the country. In a survey, 87 to 94 percent of obstetricians told Lomas that they "agreed with the content" of the guideline; 33 percent said that, as a result of the guideline, they had changed their practice of medicine. But in a follow-up test only 67 percent of the obstetricians in the survey understood the guideline's contents. The actual repeat C-section rate was 15 to 49 percent above the rate reported by the obstetricians. Lomas concluded that the guideline had produced only "slight change in actual practice." Other investigators, upon evaluating the impact of dissemination for other guidelines, have found similar results (Kosecoff et al. 1987; Merz 1991; Cohen et al. 1992).

The LDS Hospital ARDS team generated data through which members of the team could objectively evaluate their performance against the guideline. In the face of credible clinical data, in a supportive environment, guideline compliance changed. Several other projects within IHC have shown the same effect, and other investigators have reported similar results (Caper 1991).

Subjective data works no better for guideline implementation than for care delivery. Successful guideline implementation appears to rest upon the availability of an adequate data system. But such a data system serves many other concurrent purposes in addition to helping establish guideline compliance. It provides information to assess outcomes and systematically improve, and generates reports for use with regulators and health care purchasers.

Lesson 9: Physicians will lead guideline implementation if the subject is approached through existing professional values, structures, and realities.

The values and standards of the American medical profession are a ready foundation upon which a successful guideline implementation program can securely rest. But to take advantage of that foundation, administrators must approach guideline implementation from the medical perspective, using structures and language that physicians understand. For example, physicians already have a structure to implement clinical management - it is called the medical staff. Therefore, as the LDS pulmonary research team began to implement their ventilator guideline, they did not add another layer of meetings. Instead, they used their existing clinical staff meetings. Similarly, time is a limited commodity for-most practicing physicians. Asking community physicians to attend team meetings, outside of their existing quality structure, is tantamount to asking hospital employees to attend quality team meetings without compensation outside of regular work hours. At IHC, we therefore involve physicians in a supervisory role. Hospital employees invest the hours of staff work necessary to implement a guideline, regularly contacting physician leaders for oversight, direction, and approval. We call meetings of the entire subspecialty medical staff only after the staff work for a guideline or practice analysis is well advanced, and after physician leadership has already had a chance to review and criticize it.

CONCLUSION

In 1989 Linder interviewed 104 clinical and administrative leaders at 31 hospitals in the United States. All 31 hospitals used a large commercial severity-of-illness measurement system to compare the health outcomes achieved by individual physicians. All claimed to use the system as part of a quality improvement effort. But Linder concluded that 45 percent of the hospitals in the study used these tools primarily

to avoid meaningful change. Their outcomes measurement and quality improvement programs were a facade, a barrier to deflect outside criticism while they practiced "business as usual" behind their shield. For an additional 35 percent, the hospital administration used outcomes measurement and quality improvement to exert control over physicians. Only 20 percent of the hospitals surveyed used their systems to manage care processes –to build a partnership with physicians, and manage quality and cost through an informed and open discussion of difficult medical issues.

Linder's findings underscore a central issue in guideline development and implementation. Because most medical decisions have no basis in published scientific research, consensus techniques are essential to build practice guidelines. But for a significant subgroup of health care leaders, some opinions are more valuable than others. On one side is a group of "experts" (members of the guideline development team) Who can think, while on the other side (practicing clinicians) are those who can only do what they are told.

A group of "experts" can generate a practice guideline. A hospital administration can mandate that practicing physicians follow its rules. But given the realities surrounding the science of medicine, consensus methods, and the practice of medicine, effective implementation will occur only when clinicians and administrators team together to find the best patient care. That union is the sole safe haven in an increasingly competitive provider-at-risk environment. It not only creates the means to manage costs and improve patient outcomes; it generates the information necessary to market effective, efficient care to purchasers.

For the next generation of American health care systems, success will depend on the ability of health care leaders to create a culture of cooperation among all members of the health care team. Those leaders will not manage physicians. Instead, they will organize clinicians then supply them with the necessary tools, so that physicians can manage themselves and the health care processes they oversee. In creating that collaborative culture, it is obviously far more important how health care leaders implement practice guidelines, than the particular set of guidelines that they use to initiate implementation.

References

Anderson, C. A., and R. D. Daigh. "Quality Mind-Set Overcomes Barriers to Success." *Healthcare Financial Management 45* (February 1991): 21,32.

Artigas, A., J. Carlet, J. R. LeGall, C. L. Chastang, *L.* Blanch, and *R.* Fernandez. "Clinical Presentation, Prognostic Factors, and Outcomes of ARDS in the European Collaborative Study (1985-1987). In *Adult Respiratory*

Distress Syndrome, edited by W. Zapol and F. Lemaire. New York: Marcel Dekker, Inc., 1991.

Audet, A. M., S. Greenfield, and M. Field. "Medical Practice Guidelines: Current Activities and Future Directions." *Annuals of Internal Medicine* 113 (1 November 1990): 709-14.

Baird, M. L., S. W. Busboom, B. Ingram, B. C. James, S. W. Lewis, D. S. Thompson, A. Ward, and M. Weed. *Final Analysis of the IHC "Uncomplicated Total Hip Arthroplasty" Quality, Utilization, and Efficiency (QUE) Study*. Salt Lake City, UT: Intermountain Health Care, Inc., 1988.

Baird, M. L., S. W. Busboom, T. K. French, B. Ingram, B. C. James, S. W. Lewis, D. S. Thompson, A. Ward, and M. Weed. *Final Analysis of the IHC 1987 "Permanent Pacemaker Implant" Quality, Utilization, and Efficiency (QUE) Study*. Salt Lake City, UT: Intermountain Health Care, Inc., 1989.

Berwick, D. M. "TQM: Refining Doctoring." *The Internist* 34 (March 1993): 8-10.

Brook, R. H., R. E. Park, M. R. Chassin, D. H. Solomon, J. Keesey, and J. Kosecoff. "Predicting the Appropriate Use of Carotic Endarterectomy, Upper Gastrointestinal Endoscopy, and Coronary Angiography." *New England Journal of Medicine* 323 (25 October 1990): 1173-77.

Bunker, J. P. "Is Efficacy the Gold Standard for Quality Assessment?" *Inquiry* 25 (Spring 1988):51-58.

Caper, P. "Population-Based Measures of the Quality of Medical Care." In *Health Care Quality Management for the 21st Century* edited by J. Conch. Tampa, FL. American College of Physician Executives, 1991.

Cohen, M. M., N. P. Roos, L. MacWilliam, and A. Wajda. "Assessing Physicians' Compliance with Guidelines for Papanicolaou Testing." *Medical Care* 30 (June 1992): 514-28..

Deming, W. E. Personal communication. June 1990.

Dranove, D., M. Shanley, and W. D. White. "How Fast Are Hospital Prices Really Rising?" *Medical Care* 29, no. 8 (August 1991): 690-96.

Dubinsky, M., and J. H. Ferguson. "Analysis of the National Institutes of Health Medicare Coverage Assessment." *International Journal of Technology Assessment in Health Care* 6, no. 3 (1990): 480-88.

East, T. D., A. H. Morris, C. J. Wallace, T. P. Clemmer, J. F. Orme, Jr., L. K. Weaver, S. Henderson, and D. F. Sittig. "A Strategy for Development of Computerized Critical Care Decision Support Systems." *International Journal of Clinical Monitoring and Computing* 8, no. 4 (1992a): 263-69.

East, T. D., S. H. Bohm, J. Wallace, T. P. Clemmer, L. K. Weaver, J. F. Orme, and A. H. Moms. "A Successful Computerized Protocol for Clinical Management of Pressure Control Inverse Ratio Ventilation in ARDS Patients." *Chest* 101 (March 1992b): 697-710.

Eddy, D. M. The "Clinical Decision Making: From Theory to Practice" series in the *Journal of the American Medical Association*: Volumes 263-268 (1990-1992).

1. "The Challenge." *Journal of the American Medical Association* 263, no. 2 (12 Jan. 1990): 287-90.
2. "Anatomy of a Decision." *Journal of the American Medical Association* 263, no. 3 (19 Jan. 1990): 441-43.
3. "Practice Policies-What Are They*?" Journal of the American Medical Association* 263, no. 6 (9 Feb. 1990): 877-78, 880.

4. "Practice Policies: Where Do They Come From?" *Journal of the American Medical Association* 263, no. 9 (2 March 1990): 1265, 1269, 1272.
5. "Practice Policies-Guidelines for Methods." *Journal of the American Medical Association* 263, no. 13 (4 April 1990): 1839-41.
6. "Guidelines for Policy Statements: The Explicit Approach." *Journal of the American Medical Association* 263, no. 16 (25 April 1990): 223940.
7. "Comparing Benefits and Harms: The Balance Sheet." *Journal of the American Medical Association* 263, no. 18 (9 May 1990): 2493, 2498, 2501.
8. "Designing a Practice Policy: Standards, Guidelines, and Options." *Journal of the American Medical Association* 263, no. 22 (13 June 1990):3077,3081,3084.
9. "Resolving Conflicts in Practice Policies." *Journal of the American Medical Association* 264, no. 3 (18 July 1990):389-9 1:
10. "What Do We Do about Costs?" *Journal of the American Medical Association* 264, no. 9 (5 Sept. 1990): 1161, 1165, 1169-70.
11. "Connecting Value and Costs: Whom Do We Ask, and What Do We Ask Them?" *Journal of the American Medical Association 264, no.* 13 (3 Oct. 1990):1737-39.
12. "Rationing by Patient Choice." *Journal of the American Medical Association* 265, no. 1 (2 Jan. 1991): 105-8.
13. "What Care Is 'Essential'? What Services Are 'Basic'?" *Journal of the American Medical Association* 265, no. 6 (13 Feb. 1991): 782, 786-88.
14. "The Individual vs. Society: Is There a Conflict?" *Journal of the American Medical Association* 265, no. 11 (20 March 1991): 1446, 1449-50.
15. "The Individual vs. Society: Resolving the Conflict." *Journal of the American Medical Association* 265, no. 18 (8 May 1991): 2399-2401, 2405-6.
16. "What's Going on in Oregon?" *Journal of the American Medical Association* 266, no. 3 (17 July 1991):417-20.
17. "Oregon's Methods: Did Cost-Effectiveness Analysis Fail?" *Journal of the American Medical Association* 266, no. 15 (16 Oct. 1991):2135-41.
18. "Oregon's Plan: Should It Be Approved?" *Journal of the American Medical Association* 266, no. 17 (17 Nov. 1991): 2439-45.
19. "Cost-Effectiveness Analysis: A Conversation with my Father." *Journal of the American Medical Association 267, no.* 12 (25 March 1992):1669-75.
20. "Cost-Effectiveness Analysis: Is It up to the Task?" *Journal of the American Medical Association* 267, no. 24 (24 June 1992): 3342-48.
21. "Cost-Effectiveness Analysis: Will It Be Accepted?" *Journal of the American Medical Association* 268, no. 1 (1 July 1992): 132-36.
22. "Applying Cost-Effectiveness Analysis: The Inside *Story."* *Journal of the American Medical Association* 268, no. 18 (11 Nov. 1992): 2575-82.

———. *A Manual for Assessing Health Practices & Designing Practice Policies.* Philadelphia, PA: The American College of Physicians, 1992.

———. "Variations in Physician Practice: The Role of Uncertainty." *Health Affairs 3* (Summer 1984): 74-89.

Eddy, D. M., and J. Billings. "The Quality of Medical Evidence: Implications for Quality of Care." *Health Affairs 7* (Spring *1988): 19-32.*

Eddy, D. M., V. Hasselblad, and R. Shachter. *Meta-Analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence.* San Diego, CA: The Academic Press, *1992.*

Ferguson, J. H. "Research on the Delivery of Medical Care Using Hospital *Firms." Medical Care 29,* no. *7* (Supplement *1991):* JS I-2.

Flanagin, A., and G. D. Lundberg. "Clinical Decision Making: Promoting the Jump From Theory to Practice." (Editorial laying out the start of the Eddy series) *Journal of the American Medical Association 263,* no. *2 (12* Jan. *1990):279-80.*

Glover, J. A. "The Incidence of Tonsillectomy in School Children." *Proceedings of the Royal Society of Medicine 31 (1938): 1219-36.*

Goldschmidt, P. G. "Information Synthesis: A Practical Guide." *Health Services Research 21* (June, Part 1 *1986): 215-37.*

Hadorn, D. C., K. McCormick, and A. Diokno. "An Annotated Algorithm Approach to Clinical Guideline Development." *Journal of the American Medical Association 267 (24* June *1992): 3311-14.*

Henderson, S. E., R. O. Crapo, C. J. Wallace, T. D. East, A. H. Morris, and R. M. Gardner. "Performance of Computerized Protocols for the Management of Arterial Oxygenation in an Intensive Care Unit." *International Journal of Clinical Monitoring and Computing 8,* no. *4 (1992): 271-80.*

Henderson, S. E., R. O. Crapo, T. D. East, A. H. Moms, C. J. Wallace, and R. M. Gardner. "Computerized Clinical Protocols in an Intensive Care Unit: How Well Are They Followed?" *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care (SCAMC), Los* Alamitos, CA: IEEE Computer Society Press, *1990.*

Hillman, B. J., G. T. Olson, P. E. Griffith, J. H. Sunshine, C. A Joseph, S. D. Kennedy, W. R. Nelson, and L. B. Bernhardt. "Physicians' Utilization and Charges for Outpatient Diagnostic Imaging in a Medicare Population." *Journal of the American Medical Association 268 (21* October *1992): 2050-54.*

Institute of Medicine. *Assessing Medical Technologies.* Washington, DC: National Academy Press, *1985.*

——. *Clinical Practice Guidelines: Directions for a New Program.* Wash ington, DC: National Academy Press, *1990.*

James, B. C. "Good Enough? Standards and Measurement in Continuous Quality Improvement." In *Bridging the Gap Between Theory and Practice.* Chicago, IL: Hospital Research and Education Trust (American Hospital Association), 1992.

——. *Quality Management for Health Care Delivery.* Chicago, IL: Hospital Research and Educational Trust (American Hospital Association), *1989.*

James, B. C., M. Baird, A. Ward, M. Weed, S. W. Lewis, and S. Busboom. *Final Analysis of the IHC "Uncomplicated Cholecystectomy" Quality. Utilization, and Efficiency (QUE) Study: Intermountain Health Care Department of Medical Affairs Technical Report No. 2.* Salt Lake City, UT: Intermountain Health Care, Inc., 1988.

James, B. C., M. Weed, S. W. Lewis, S. Busboom, G. Darricades, and B Ingram. *Final Analysis of the IHC "Transurethral Prostatectomy"*

Utilization Study: Intermountain Health Care Department of Medical Affairs Technical Report No. 1. Salt Lake City, UT: Intermountain Health Care, Inc., 1987.

James, B. C., S. D. Horn, and R. A. Stephenson. Management by Fact: The Relationship of Quality Improvement to Outcomes Management, Practice Guidelines, and Randomized Clinical Trials. Unpublished manuscript, 1992.

Kellie, S. E., and J. T. Kelly. "Medicare Peer Review Organization Preprocedure Review Criteria." *Journal of the American Medical Association* 265 (13 March 1991): 1265-70.

Koran, L. M. "The Reliability of Clinical Methods, Data and Judgments." *New England Journal of Medicine* 293 (25 September 1975 (Part 1), 2 October 1975 (Part 2)): 642-46; 695-701.

Kosecoff, J., D. E. Kanouse, W. H. Rogers, L. McCloskey, C. M. Winslow, and R. H. Brook. "Effects of the National Institutes of Health Consensus Development Program on Clinical Practice." *Journal of the American Medical Association* 258 (20 November 1987): 2708-13.

Leape, L. L., R. E. Park, D. H. Solomon, M. R. Chassin, J. Kosecoff, and R. H. Brook. "Does Inappropriate Use Explain Small-Area Variation in the Use of Health Care Services?" *Journal of the American Medical Association* 263 (2 February 1990): *669-72.*

Leape, L. L., R. E. Park, J. P. Kahan, and R. H. Brook. "Group Judgments of Appropriateness: The Effect of Panel Composition." *Quality Assurance in Health Care 4,* no. 2 (1992): 151-59.

Lewis, C. E. "Variations in the Incidence of Surgery." *New England Journal of Medicine* 281, no. 16 (1969): 880-84.

Linder, J. C. "Outcomes Measurement: Compliance Tool or Strategic Initia tive?" *Health Care Management Review 16* (Fall 1991): 21-33.

Lomas, J., G. M. Anderson, K. Domnick-Pierre, E. Vayda, M. W. Enkin, and W. J. Hannah. "Do Practice Guidelines Guide Practice?: The Effect of a Consensus Statement on the Practice of Physicians." *New England Journal of Medicine* 321 (9 November 1989): 1306-11.

McDonald, C. J. "Protocol-Based Computer Reminders, the Quality of Care and the Non-Perfectibility of Man." *New England Journal of Medicine* 295 (9 December 1976): 1352-55.

McPherson, K., J. E. Wennberg, O. B. Hovind, and P. Clifford. "Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway." *New England Journal of Medicine* 307 (18 November 1982): 1310-14.

Merz, B. "High Consensus, Low Compliance on Cancer Screening Guidelines." *American Medical News* September 23/30,1991.

Mitchell, J. M., and E. Scott. "Physician Ownership of Physical Therapy Services: Effects on Charges, Utilization, Profits, and Service Characteristics." *Journal of the American Medical Association 268 (21* October 1992): 2055-59.

Mitchell, J. M., and J. H. Sunshine. "Consequences of Physicians' Ownership of Health Care Facilities-Joint Ventures in Radiation Therapy." *New England Journal of Medicine 327 (19* November 1992):1497-1501.

Morris, A. H. "Computer Applications." In *Principles of Critical Care,* edited by J. B. Hall, G. A. Schmidt, and L. D. H. Wood. New York: McGraw-Hill, Inc., Health Professions Division, 1992.

——— . "Protocols, ECCO2R, and the Evaluations of New ARDS Therapy." *Japanese Journal of Intensive Care Medicine* 16 (1992): 61-63.

Morris, A. H., C. J. Wallace, T. P. Clemmer, J. F. Orme, Jr., L. K. Weaver, F. Thomas, N. C. Dean, R. Menlove, and T. D. East. "Final Report: Computerized Protocol Controlled Clinical Trial of New Therapy which Includes ECCO2R for ARDS." *American Review Respiratory Disease* 145, no. 4 (1992): A 184.

O'Connor, G. T., S. Plume, J. Beck, C. Martin, W. Nugent, and E. Olmstead. "What Are My Chances? It Depends on Whom You Ask. The Choice of a Prosthetic Heart Value." *Medical Decision Making* 18 (October-December 1988): 341.

Park, R. E., A. Fink, R. H. Brook, M. Chassin, K. Kahn, N. Merrick, J. Kosecoff, and D. Solomon. *Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures.* R-3280-CWF/HF/PMT/RWJ. Santa Monica, CA: RAND Corporation, 1986.

——— . "Physician Ratings of Appropriate Indications for Three Procedures: Theoretical indications vs. Indications Used in Practice." *American Journal of Public Health* 79 (April 1989): 445-47.

Pearson, S. D., C. Z. Margolis, S. Davis, L. K. Schreier, and L. K. Gottlieb. "The Clinical Algorithm Nosology: A Method for Comparing Algorithmic Guidelines." *Medical Decision Making* 12 (April-June 1992): 123-31.

Ryan, T. P. *Statistical Methods for Quality Improvement.* New York, NY: John Wiley and Sons, 1989.

Scherkenbach, W.W.*The Deming Route to Quality and Productivity: Road Maps and Roadblocks.* Washington, DC: CEEPress Books, 1991.

Society for Medical Decision Making Committee on Standardization of Clinical Algorithms. "Proposal for Clinical Algorithm Standards." *Medi cal Decision Making 12* (April-June 1992): 149-54.

Swedlow, A., G. Johnson, N. Smithline, and A. Milstein. "Increased Costs and Rates of Use in the California Workers' Compensation System as a Result of Self-Referral by Physicians." *New England Journal of Medicine* 327 (19 November 1992): 1502-6.

Wennberg, J. E. "Variations in Medical Practice and Hospital Costs." *Connecticut Medicine 49,* no. 7 (1985): 411 53.

Wennberg, J. E., and A. Gittelsohn. "Small Area Variations in Health Care Delivery." *Science* 182 (14 December 1973): 1102-8.

Wennberg, J. E., B. A. Barnes, and M. Zubkoff. "Professional Uncertainty and the Problem of Supplier-Induced Demand." *Social Science and Medicine* 16, no. 7 (1982): 811-24.

Wennberg, J. E., L. Blowers, R. Parker, and A. M. Gittelsohn. "Changes in Tonsillectomy Rates Associated with Feedback and Review." *Pediatrics* 59 (June 1977): 821-86.

Williamson, J. W., M. Alexander, and G. E. Miller. "Priorities in Patient Care Research and Continuing Medical Education." *Journal of the American Medical Association* 204 (22 April 1968): 303-8.

Williamson, J. W., M. Lincoln, and C. W. Turner. Personal communication, May 1991.

Williamson, J. W., P. G. Goldschmidt, and 1. A. Jillson. *Medical Practice Information Demonstration Project: Final Report.* Office of the Asst. Secretary of Health, DHEW, Contract #282-77-0068GS. Baltimore, MD: Policy Research Inc., 1979.

Winslow, C. M., J. Kosecoff, M. R. Chassin, D. E. Kanouse, and R. H. Brook. "The Appropriateness of Performing Coronary Artery Bypass Surgery." *Journal of the American Medical Association 260* (22 July 1988): 505-9.

Zapol, W. M., M. J. Frikker, H. Pontoppidan, R. S. Wilson, and K. E. Lynch. "The Adult Respiratory Distress Syndrome at Massachusetts General Hospital: Etiology Progression and Survival Rates, 1978-1988." In *Adult Respiratory Distress Syndrome,* edited by F. Walz. New York: Marcel Dekker, Inc., 1991.

Used with permission from Frontiers of Health Service Management, Volume 10:1: 3-37. (Chicago: Health Administration Press, Fall, 1993)